Title

Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data

Journal

Journal of Geophysical Research, 119(12)

ISSN

0148-0227

Authors

Wecht, KJ
Jacob, DJ
Frankenberg, C
et al.

Publication Date

2014-06-27

DOI

10.1002/2014JD021551

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2014JD021551

Key Points:
- Satellite observations are used to estimate methane emissions from North America
- We demonstrate a new method for evaluating information content of observations
- U.S. methane emissions consistent with EPA estimate, but livestock higher by 40%

Correspondence to:
K. J. Wecht,
wecht@fas.harvard.edu

Citation:
Wecht, K. J., D. J. Jacob, C. Frankenberg, Z. Jiang, and D. R. Blake (2014), Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data, *J. Geophys. Res. Atmos.*, *119*, 7741–7756, doi:10.1002/2014JD021551.

Received 27 JAN 2014
Accepted 6 MAY 2014
Accepted article online 9 MAY 2014
Published online 26 JUN 2014

Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data

Kevin J. Wecht[1], Daniel J. Jacob[1], Christian Frankenberg[2], Zhe Jiang[2], and Donald R. Blake[3]

[1]Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts, USA, [2]Jet Propulsion Laboratory/California Institute of Technology, Pasadena, California, USA, [3]Department of Chemistry, University of California, Irvine, California, USA

Abstract We estimate methane emissions from North America with high spatial resolution by inversion of Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) satellite observations using the Goddard Earth Observing System Chemistry (GEOS-Chem) chemical transport model and its adjoint. The inversion focuses on summer 2004 when data from the Intercontinental Chemical Transport Experiment-North America (INTEX-A) aircraft campaign over the eastern U.S. are available to validate the SCIAMACHY retrievals and evaluate the inversion. From the INTEX-A data we identify and correct a water vapor-dependent bias in the SCIAMACHY data. We conduct an initial inversion of emissions on the horizontal grid of GEOS-Chem (1/2° × 2/3°) to identify correction tendencies relative to the Emission Database for Global Atmospheric Research (EDGAR) v4.2 emission inventory used as a priori. We then cluster these grid cells with a hierarchical algorithm to extract the maximum information from the SCIAMACHY observations. A 1000 cluster ensemble can be adequately constrained, providing ~100 km resolution across North America. Analysis of results indicates that the Hudson Bay Lowland wetlands source is 2.1 Tg a^{-1}, lower than the a priori but consistent with other recent estimates. Anthropogenic U.S. emissions are 30.1 ± 1.3 Tg a^{-1}, compared to 25.8 Tg a^{-1} and 28.3 Tg a^{-1} in the EDGAR v4.2 and Environmental Protection Agency (EPA) inventories, respectively. We find that U.S. livestock emissions are 40% greater than in these two inventories. No such discrepancy is apparent for overall U.S. oil and gas emissions, although this may reflect some compensation between overestimate of emissions from storage/distribution and underestimate from production. We find that U.S. livestock emissions are 70% greater than the oil and gas emissions, in contrast to the EDGAR v4.2 and EPA inventories where these two sources are of comparable magnitude.

1. Introduction

Methane is the second most important anthropogenic greenhouse gas after carbon dioxide [*Myhre et al.*, 2013]. Major anthropogenic sources include natural gas extraction and use, coal mining, landfills, livestock, rice cultivation, and biomass and biofuel burning. Wetlands are the largest natural source. The magnitude of global methane emissions is constrained within ±15% by knowledge of the global sink from oxidation by OH, but the magnitudes and trends of emissions from different source types and source regions are highly uncertain [*Myhre et al.*, 2013; *Hartmann et al.*, 2013; *Kirschke et al.*, 2013]. Reducing methane emissions has been identified as a low-cost priority in greenhouse gas emissions reduction strategies [*International Energy Agency*, 2013; *van Vuuren et al.*, 2010; *Weyant et al.*, 2006], but this requires that the sources be quantified. The United States (U.S.) Environmental Protection Agency (EPA) provides national emission inventories for methane [*United States Environmental Protection Agency (EPA)*, 2013]. However, a number of studies using atmospheric observations from surface and aircraft suggest that these inventories may underestimate total emissions or emissions from various source types by a factor of two or more [*Katzenstein et al.*, 2003; *Xiao et al.*, 2008; *Kort et al.*, 2008; *Pétron et al.*, 2012; *Miller et al.*, 2013; *Karion et al.*, 2013; G. W. Santoni, submitted manuscript, 2014].

Satellite observations of atmospheric methane provide a resource for constraining emissions, as first demonstrated by *Bergamaschi et al.* [2007]. Satellites deliver dense spatial coverage unachievable by surface networks or aircraft campaigns, albeit with lower precision. Methane has been retrieved from nadir satellite measurements of solar backscatter in the short-wave infrared (SWIR) and terrestrial radiation in the thermal infrared (TIR). SWIR retrievals are available from Scanning Imaging Absorption Spectrometer

for Atmospheric Chartography (SCIAMACHY) for 2003–2012 [*Frankenberg et al.*, 2011] and Greenhouse Gases Observing Satellite (GOSAT) for 2009 to present [*Parker et al.*, 2011; *Schepers et al.*, 2012]. TIR retrievals are available from Atmospheric Infrared Sounder for 2002 to present [*Xiong et al.*, 2008], Tropospheric Emission Spectrometer for 2004–2011 [*Worden et al.*, 2012], and Infrared Atmospheric Sounding Interferometer for 2007 to present [*Xiong et al.*, 2013; *Crevoisier et al.*, 2013]. SWIR retrievals provide total atmospheric columns. TIR retrievals provide vertical profiles but with low sensitivity to the lower troposphere due to lack of thermal contrast, and this limits their value for detecting regional sources [*Wecht et al.*, 2012]. SCIAMACHY had full global coverage with a 6 day return time. Current coverage by GOSAT is much sparser. Instrument degradation limited the value of the SCIAMACHY data after 2005 [*Frankenberg et al.*, 2011]. The Topospheric Monitoring Instrument to be launched in 2015 will provide SWIR methane data with global daily coverage and $7 \times 7\,\mathrm{km}^2$ nadir resolution [*Veefkind et al.*, 2012].

Here we use SCIAMACHY observations for July–August 2004 in an inversion of methane sources in North America with the adjoint of the Goddard Earth Observing System Chemistry (GEOS-Chem) chemical transport model (CTM) at $1/2° \times 2/3°$ (~$50 \times \mathrm{km}^2$) resolution. This time window takes advantage of concurrent methane observations from the NASA Intercontinental Chemical Transport Experiment-North America (INTEX-A) aircraft mission over the eastern U.S. [*Singh et al.*, 2006] that offer extensive vertical profile information (for satellite validation) and boundary layer mapping (for complementary source characterization). The *EPA* [2013] emission inventory shows no significant change from 2005 to 2011, implying that there is little interannual variability in U.S. anthropogenic emissions and that constraints on 2004 emissions should therefore be relevant to present day.

A number of previous studies have used SCIAMACHY data for global inverse modeling of methane sources [*Bergamaschi et al.*, 2007, 2009, 2013; *Meirink et al.*, 2008; *Monteil et al.*, 2013; *Cressot et al.*, 2014; *Houweling et al.*, 2013]. All have recognized the need for correcting bias in the SCIAMACHY data that otherwise propagate to the inverse solution. An early validation of SCIAMACHY using ground based Fourier transform spectrometers [*Dils et al.*, 2006] failed to identify retrieval error related to inaccuracies in water vapor spectroscopic parameters [*Frankenberg et al.*, 2008]. More recently, *Houweling et al.* [2013] show that bias in SCIAMACHY is correlated with tropospheric water vapor concentrations. As we show below, a water vapor correction enables successful validation of the SCIAMACHY data with the INTEX-A vertical profiles.

Our work goes beyond the above studies in using SCIAMACHY for a continental-scale optimization of methane sources with high resolution, including evaluation with aircraft data. The adjoint-based approach allows us to exploit the density of the satellite observations to optimize emissions at the $1/2° \times 2/3°$ native resolution of GEOS-Chem, but we show that too fine a resolution can inhibit successful inversion by diluting the information from the observations. Previous studies have proposed methods for coarsening the discretization of model emissions in a way that optimizes the inversion [*Bocquet,* 2005*,* 2009; *Bocquet et al.*, 2011; *Wu et al.*, 2011]. These methods require computationally intensive construction of the Jacobian of the CTM, which is precisely what we seek to avoid by using the adjoint method. Here we introduce a hierarchical clustering algorithm to optimize the discretization of emissions in the context of adjoint-based inverse modeling.

2. Observations

SCIAMACHY is in a Sun-synchronous polar orbit with an equator overpass local time of ~10:00. It retrieves methane from nadir SWIR spectra at 1.66–1.67 μm with a nadir footprint of $30 \times 60\,\mathrm{km}^2$ and cross-track scanning. It achieves complete global coverage every 6 days. Observations are limited to daytime and land. We use the Iterative Maximum A Posteriori (IMAP) v5.5 retrieval from *Frankenberg et al.* [2011]. The retrieval first calculates the methane vertical column density Ω_{CH4} (molecules cm^{-2}):

$$\Omega_{CH4} = \Omega_{\mathbf{A}} + \mathbf{a}^T(\omega - \omega_{\mathbf{A}}) \quad (1)$$

where ω is the true vertical profile of methane, consisting of 20 partial columns (molecules cm^{-2}) on a vertical grid, $\omega_{\mathbf{A}}$ is the a priori profile provided by the TM5-4DVAR CTM [*Meirink et al.*, 2008], $\Omega_{\mathbf{A}}$ is the corresponding a priori column concentration, and **a** is an averaging kernel vector that describes the sensitivity of the retrieved column to each partial column in ω. The sensitivity measured by **a** is nearly uniform throughout the troposphere and decreases with altitude in the stratosphere. To account for the impact of aerosols and partial

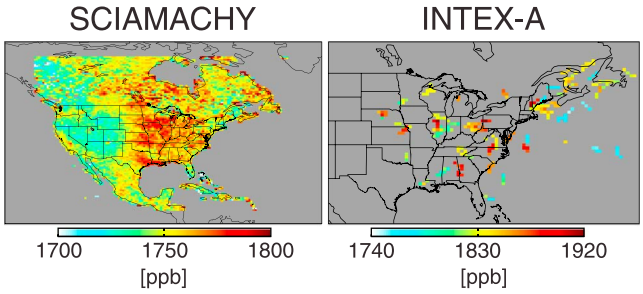

Figure 1. Methane over North America during the INTEX-A aircraft campaign (1 July to 14 August 2004): (left) SCIAMACHY mean column mixing ratios and (right) INTEX-A mixing ratios below 850 hPa. Observations are averaged spatially on the 1/2° × 2/3° GEOS-Chem grid and temporally over the INTEX-A campaign duration. Note the difference in scales between the two panels.

cloud cover on the observed light path, Ω_{CH4} is normalized and converted to a column mixing ratio X_{CH4} (mol mol^{-1}) using the CO_2 proxy method described in detail by *Frankenberg et al.* [2006]:

$$X_{CH4} = (\Omega_{CH4}/\Omega_{CO2})X_{CO2} \quad (2)$$

where Ω_{CO2} is the vertical column density of CO_2 also retrieved by SCIAMACHY, and X_{CO2} is a modeled column mixing ratio of CO_2. CO_2 is used for normalization because it is retrieved in a spectrally neighboring fitting window, and its mixing ratio is known with much higher precision than methane.

The IMAP v5.5 product was previously validated by *Houweling et al.* [2013], who used coincident observations from the Total Carbon Column Observing Network to identify a seasonally dependent bias that they attributed to water vapor. Here we use in situ vertical profiles from the INTEX-A aircraft during summer 2004. The aircraft flew over the eastern U.S. with extensive boundary layer coverage (Figure 1, right) and vertical profiles extending up to 12 km. Methane was measured using gas chromatography from whole air flask samples collected every 4 min with accuracy of 1.0 ppb and precision of 0.1 ppb [*Colman et al.*, 2001] (http://www-air.larc.nasa.gov/cgi-bin/arcstat). For SCIAMACHY validation we require vertical profiles that span from at least 900 to 400 hPa and coincide with SCIAMACHY overpasses within ± 150 km and ±6 h. We find nine profiles satisfying these criteria, each corresponding to 7 to 29 satellite observations. Further tightening the spatiotemporal requirements would exclude all INTEX-A profiles. We map the aircraft profiles on the 12 levels of the SCIAMACHY retrieval pressure grid, extrapolate above the DC-8 ceiling using the SCIAMACHY a priori profile, and apply equation (1) to simulate the SCIAMACHY retrieval. From there we derive X_{CH4} by dividing by the local air column density. Because the SCIAMACHY retrieval uses modeled X_{CO2}, we use the same modeled X_{CO2} rather than INTEX-A observed X_{CO2}. We average the coincident SCIAMACHY observations and compute the SCIAMACHY-INTEX difference ΔX_{CH4}. Results indicate a mean bias $\Delta X_{CH4} = -14.2$ ppb (0.8%) and a residual standard deviation of 29.2 ppb (1.6%) for individual SCIAMACHY observations.

Previous studies have demonstrated the need for a latitudinally dependent SCIAMACHY bias correction [*Bergamaschi et al.*, 2007, 2009, 2013; *Meirink et al.*, 2008; *Cressot et al.*, 2014]. Some have documented the interference of water vapor as the cause of the bias [*Frankenberg et al.*, 2008; *Houweling et al.*, 2013] and we seek such a relationship here. Figure 2 (left) shows the relationship of ΔX_{CH4} with the average pressure-weighted specific humidity in the 900–400 hPa column measured by the INTEX-A aircraft. There is a linear relationship

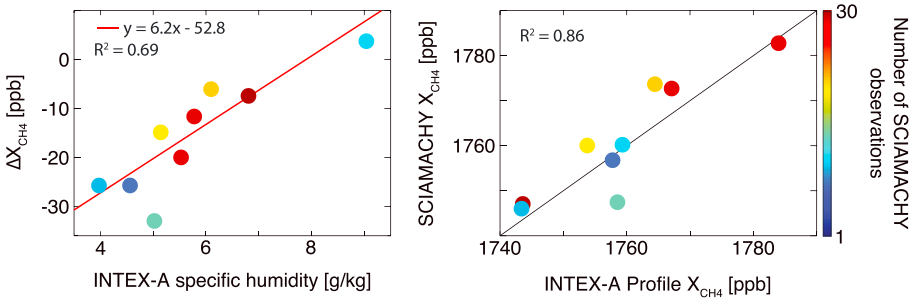

Figure 2. Validation of the SCIAMACHY IMAP v5.5 retrieval of methane column mixing ratio (X_{CH4}) with coincident INTEX-A aircraft vertical profiles (see text). (left) Difference ΔX_{CH4} between SCIAMACHY and INTEX-A plotted as a function of the mean pressure-weighted specific humidity in the 900–400 hPa column measured by the INTEX-A aircraft. The red line shows a linear regression weighted by the number of SCIAMACHY observations. Regression parameters and weighted R^2 are shown inset. (right) Comparison of SCIAMACHY and INTEX-A X_{CH4} after applying the water vapor correction from the linear regression. Weighted R^2 is shown inset, and the 1:1 line is also shown. Colors represent the number of SCIAMACHY observations averaged around each INTEX-A profile.

(weighted $R^2 = 0.69$) that implies a negative bias under dry conditions and a positive dependence of the bias on humidity. We use this relationship to calculate a linear bias correction factor and apply it to the original IMAP v5.5 retrieval. After the correction we find a negligibly small mean bias $\Delta X_{CH4} = 2.5$ ppb (0.1%) and residual standard deviation of 28.2 ppb (1.6%) for individual SCIAMACHY observations (Figure 2, right). The residual standard deviation (which we take to represent SCIAMACHY random measurement error) is consistent with the average IMAP v5.5 theoretical error of 30.9 ppb (1.7%) reported by *Frankenberg et al.* [2011]. All SCIAMACHY data shown here include the specific humidity correction applied with the local GEOS-5 meteorological data used to drive GEOS-Chem.

Figure 1 (left) shows mean SCIAMACHY methane column mixing ratios during the INTEX-A period. Values are highest over the central U.S. where there are large sources from livestock and from natural gas and oil (oil and gas) production. Values are also high over the Canadian wetlands in northern Ontario. The low values in the west reflect elevated terrain so that the stratosphere (where methane is depleted) makes a relatively large contribution to the column mixing ratio. Also, shown in Figure 1 are the individual INTEX-A observations in the boundary layer (below 850 hPa). These show areas of high concentrations in the Midwest and east but with fine-scale structure that must reflect in part day-to-day variability in meteorology. We do not use the INTEX-A data for the inversion but use them instead for validation (as described above) and evaluation of the inversion results obtained from SCIAMACHY.

3. Optimization of Methane Emissions

We optimize methane emissions in North America on the basis of SCIAMACHY observations by Bayesian inference, adjusting an initial (a priori) emission field in order to minimize the difference in X_{CH4} between SCIAMACHY and GEOS-Chem with error weighting. We describe the procedure and its evaluation below.

3.1. GEOS-Chem Model and A Priori Emissions

We use the GEOS-Chem CTM v9-01-02 (http://acmg.seas.harvard.edu/geos/index.html) as the forward model for the inversion. GEOS-Chem is driven by GEOS-5 meteorological data from the NASA Global Modeling and Assimilation Office. The GEOS-5 data have 1/2° latitude × 2/3° longitude horizontal resolution and 6 h temporal resolution (3 h for surface variables and mixing depths). Here we use the native 1/2° × 2/3° resolution for GEOS-Chem over North America and adjacent oceans (10–70°N, 40–140°W), with 3 h dynamic boundary conditions from a global simulation with 4° × 5° resolution. This nested North American functionality of GEOS-Chem has been used previously in a number of air quality studies including extensive evaluation with observations [*Park et al.*, 2004, 2006; *L. Zhang et al.*, 2011, 2012; *Y. Zhang et al.*, 2012; *van Donkelaar et al.*, 2012]. These show a good simulation of regional transport with no apparent biases.

The GEOS-Chem methane simulation was originally described by *Wang et al.* [2004] and updated by *Pickett-Heaps et al.* [2011]. The main methane sink is tropospheric oxidation by OH, computed using a 3-D archive of monthly average OH concentrations from a GEOS-Chem simulation of tropospheric chemistry [*Park et al.*, 2004]. The mean mass-weighted tropospheric OH concentration is 10.8×10^5 molecules cm^{-3}. Additional minor sinks for methane are soil absorption [from *Fung et al.*, 1991] and oxidation in the stratosphere. We use stratospheric methane loss frequencies archived from the NASA Global Modeling Initiative model [*Considine et al.*, 2008; *Allen et al.*, 2010] as described by *Murray et al.* [2012]. The resulting global mean atmospheric lifetime of methane is 8.9 years, and the lifetime against oxidation by tropospheric OH is 9.9 years. Model intercomparisons in the literature give corresponding values of 8.6 ± 1.2 years and 9.8 ± 1.6 years [*Voulgarakis et al.*, 2013]. *Prather et al.* [2012] estimate corresponding values of 9.1 ± 0.9 years and 11.2 ± 1.3 years from observational constraints.

For the a priori emissions we use the 2004 anthropogenic inventory from Emission Database for Global Atmospheric Research (EDGAR) v4.2 with 0.1° × 0.1° resolution and no seasonality [*European Commission, Joint Research Centre/Netherlands Environmental Assessment Agency*, 2009]. Natural sources include temperature-dependent emissions from wetlands [*Kaplan*, 2002; *Pickett-Heaps et al.*, 2011], termites [*Fung et al.*, 1991], and daily Global Fire Emissions Database Version 3 open fire emissions [*van der Werf et al.*, 2010; *Mu et al.*, 2010]. Figure 3 shows total methane emissions for North America and the contributions from the five largest source types.

Table 1 lists U.S. anthropogenic emission totals by source type in the EDGAR v4.2 and EPA inventories (the EPA inventory is available only as a national total). Total U.S. anthropogenic emissions from EDGAR v4.2 and EPA are 25.8 and 28.3 Tg a^{-1}, respectively. EDGAR v4.2 and EPA give similar estimates for emissions by source type, except for oil and gas and coal mining. EDGAR reports oil and gas emissions of 6.3 Tg a^{-1},

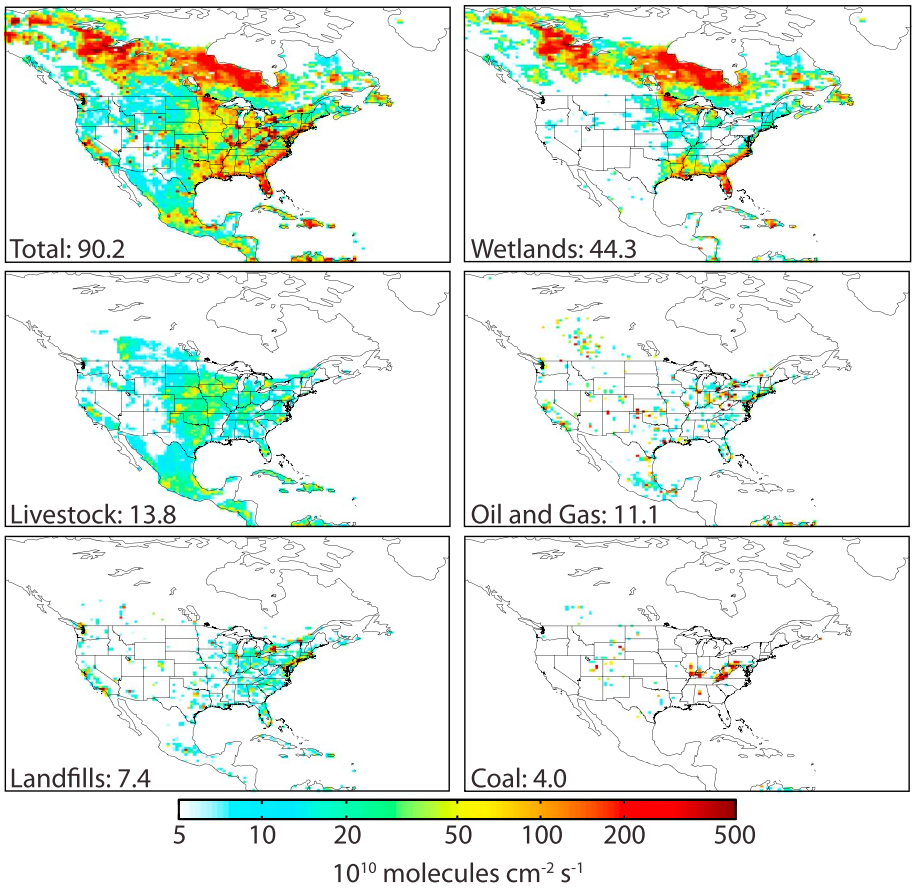

Figure 3. North American methane emissions used as a priori for the inversion: (top left) total emissions and contributions from the major source types. Inventories are from *Kaplan et al.* [2002] and *Pickett-Heaps et al.* [2011] for wetlands and from Global Emissions Inventory Activity v4.2 for all other (anthropogenic) sources. Values are averages for 22 June to 14 August 2004. Annual emission rates for 2004 (Tg a^{-1}) are shown inset for the North America domain as encompassed by the figure.

30% lower than the *EPA* [2013] estimate of 9.0 Tg a^{-1}. It reports U.S. coal mining emissions of 3.9 Tg a^{-1}, 40% higher than the *EPA* [2013] estimate of 2.7 Tg a^{-1}.

Figure 4 shows surface air methane concentrations from the global and nested GEOS-Chem simulations with a priori emissions as described above, compared to observations from the NOAA Global Monitoring Division

Table 1. U.S. Fluxes of Methane in 2004 [Tg a^{-1}]

Source Type	*EPA* [2013][a]	EDGAR v4.2[b]	*Miller et al.* [2013][c]	This Work[d]
Total			47.2 ± 1.9	37.0 ± 1.4
Anthropogenic	28.3 (24.6, 32.3)	25.8	44.5 ± 1.9	30.1 ± 1.3
Livestock	8.8 (7.7, 10.4)	8.5	16.9 ± 6.7	12.2 ± 1.3
Natural Gas and Oil	9.0 (7.2, 13.4)	6.3		7.2 ± 0.6
Landfills	5.4 (2.5, 7.9)	5.3		5.8 ± 0.3
Coal Mining	2.7 (2.3, 3.2)	3.9		2.4 ± 0.3
Other[e]	2.4 (1.4, 4.2)	1.9		2.5 ± 0.2
Natural[f]			2.7	6.9 ± 0.5

[a] Values in parentheses represent lower and upper ends of 95% uncertainty ranges.
[b] Used as a priori for the inversion.
[c] Annual average emissions for 2007–2008.
[d] Annual fluxes extrapolated from an inversion during 22 June to 14 August 2004. Uncertainties on emissions from individual source types assume that the relative a priori source distribution in each 0.5° × 0.67° grid cell is correct (see text).
[e] Including waste water treatment, rice cultivation, biofuel use, and other small sources. According to *EPA* [2013], none of these sources account for more than 3% of total U.S. anthropogenic emissions.
[f] Including natural wetlands, termites, open fires, and soil absorption.

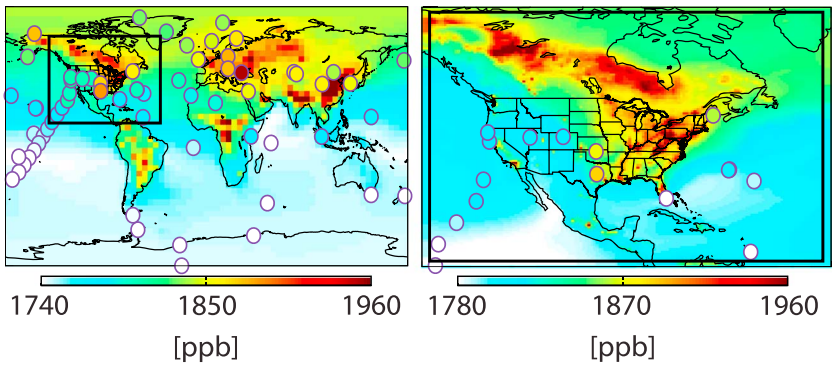

Figure 4. Methane concentrations in surface air averaged over the inversion period (22 June to 14 August 2004). The GEOS-Chem simulation with a priori sources (background) is compared to NOAA GMD observations (circles). (left) Global simulation at 4° × 5° resolution used to archive a priori boundary concentrations for the nested simulation. (right) Nested simulation at 1/2° × 2/3° resolution for the North America domain. The thick line represents the boundaries for the nested simulation. Note the difference in scale between panels. The NOAA GMD data were obtained from http://www.esrl.noaa.gov/gmd/.

network (http://www.esrl.noaa.gov/gmd/). Boundary concentrations for the nested grid are archived at the edge of the North America domain. Comparison of GEOS-Chem with the NOAA data over the remote oceans shows that the model simulates realistic latitudinal gradients, and this is further supported by comparison to High-performance Instrumented Airborne Platform for Environmental Research Pole-to-Pole Observations (HIPPO) pole-to-pole aircraft observations over the Pacific [*Wofsy et al.*, 2011; *Turner et al.*, 2013].

3.2. Inversion Method

We seek to use the SCIAMACHY observations over North America to optimize methane emissions on the 1/2° × 2/3° GEOS-Chem grid. Consider the ensemble of SCIAMACHY observations (column mean methane mixing ratios) assembled into an observation vector **y**. We assemble the gridded emissions and the gridded boundary conditions for GEOS-Chem into a state vector **x**. Let F represent GEOS-Chem serving as forward model for the inversion. We have

$$\mathbf{y} = F(\mathbf{x}) + \varepsilon \tag{3}$$

where ε is the observational error and includes contributions from forward model error, representation error (sampling mismatch between observations and the model), and measurement error. Error statistics are represented by the observational error covariance matrix $\mathbf{S_O} = E[\varepsilon\varepsilon^T]$ where $E[\,]$ is the expected value operator.

Bayesian optimization weighs the constraints on **x** from the SCIAMACHY observations with the a priori estimates $\mathbf{x_A}$ (error covariance matrix $\mathbf{S_A}$). Applying Bayes' theorem and assuming Gaussian errors leads to an optimized estimate for **x** by minimizing the cost function $J(\mathbf{x})$ [*Rodgers*, 2000]:

$$J(\mathbf{x}) = (F(\mathbf{x}) - \mathbf{y})^T \mathbf{S_O}^{-1} (F(\mathbf{x}) - \mathbf{y}) + (\mathbf{x} - \mathbf{x_a})^T \mathbf{S_A}^{-1} (\mathbf{x} - \mathbf{x_A}) \tag{4}$$

Minimization of $J(\mathbf{x})$ is done with the GEOS-Chem adjoint model, developed by *Henze et al.* [2007] and previously applied to methane source optimization by *Wecht et al.* [2012]. The adjoint calculates $\nabla_{\mathbf{x}} J(\mathbf{x_A})$, passes it to a steepest-descent algorithm that returns an improved estimate $\mathbf{x}_1$ for **x**, calculates $\nabla_{\mathbf{x}} J(\mathbf{x}_1)$, and iterates until convergence to find $\nabla_{\mathbf{x}} J(\mathbf{x}) = 0$. We describe below in more detail the different components of the inversion.

The ability of the inversion to constrain methane emissions over North America is contingent on the model variability being driven by these emissions. Starting from initial conditions, we find that it takes about a week for variability of methane columns over North America in the nested model to be driven by fresh emissions and boundary conditions (as opposed to the initial conditions). We therefore initialize our simulation on 22 June 2004, 9 days prior to assimilating the first observations on 1 July. The inversion period over which we solve for emissions is 22 June to 14 August; observations are assimilated from 1 July to 14 August. The lifetime of methane against oxidation by OH is sufficiently long to play no significant role in the variability of methane concentrations over the North America domain. Prescribed OH concentrations used in the model are therefore of no significant consequence to the inversion results.

We attempted at first to optimize North American emissions and boundary conditions as a single state vector in the inversion. This was not successful because boundary conditions have a much larger impact in determining methane concentrations, even if they are less important for determining variability. We therefore iteratively minimize two separate cost functions, $J(\mathbf{x_b})$ and $J(\mathbf{x_e})$, to optimize boundary concentrations and emissions, respectively:

$$J(\mathbf{x_B}) = (F(\mathbf{x_B}) - \mathbf{y})^T \mathbf{S_O}^{-1} (F(\mathbf{x_B}) - \mathbf{y}) + (\mathbf{x_B} - \mathbf{x_{B,A}})^T \mathbf{S_{B,A}}^{-1} (\mathbf{x_B} - \mathbf{x_{B,A}}) \quad (5)$$

$$J(\mathbf{x_E}) = (F(\mathbf{x_E}) - \mathbf{y})^T \mathbf{S_O}^{-1} (F(\mathbf{x_E}) - \mathbf{y}) + (\mathbf{x_E} - \mathbf{x_{E,A}})^T \mathbf{S_{E,A}}^{-1} (\mathbf{x_E} - \mathbf{x_{E,A}}) \quad (6)$$

Here the state vectors are $\mathbf{x_B}$, scale factors of boundary concentrations at the edge of the North American model domain relative to the a priori, and $\mathbf{x_E}$, logarithms of scale factors of methane emissions relative to the a priori within the North America domain. We optimize the logarithms of the emission scale factors to ensure positivity in the optimized emissions. A priori values for $\mathbf{x_B}$ and $\mathbf{x_E}$ are labeled $\mathbf{x_{B,A}}$ and $\mathbf{x_{E,A}}$, respectively, and the corresponding a priori error covariance matrices are $\mathbf{S_{B,A}}$ and $\mathbf{S_{E,A}}$.

Each element of $\mathbf{x_B}$ represents a temporally averaged scale factor applied to a $4° \times 5°$ grid cell along the boundaries of the North American model domain and extending over 47 vertical levels, for a total of 3290 elements. A priori boundary concentrations are specified from the global GEOS-Chem simulation with a priori emissions (shown in Figure 4). The a priori error covariance matrix $\mathbf{S_{B,A}}$ is constructed using error statistics from HIPPO-GEOS-Chem comparisons over the central Pacific presented by *Turner et al.* [2013]. The diagonal is populated with a model error standard deviation of 16 ppb (0.9%), and off-diagonal terms are parameterized with exponential error correlation length scales of 275 km in the horizontal and 78 hPa in the vertical [*Wecht et al.*, 2012]. We assume that the above error statistics apply to all four boundaries.

Each element $x_{E,i,j}$ of $\mathbf{x_E}$ represents a temporally averaged scale factor applied to total emissions from each $1/2° \times 2/3°$ emitting grid cell (i,j) in North America for a total of 7906 elements. It is expressed as follows:

$$x_{E,i,j} = \ln\left(E_{i,j}/E_{A,i,j}\right) \quad (7)$$

where $E_{i,j}$ is the true emission flux and $E_{A,i,j}$ is the a priori described above.

The a priori error covariance matrix for the emissions, $\mathbf{S_{E,A}}$, is constructed by assuming a uniform relative error standard deviation of 30% for emissions from each model grid cell and no a priori error correlations (diagonal matrix). The sensitivity of the optimized solution to the specification of a priori error will be discussed later by considering an inversion without a priori constraints.

The observational error covariance matrix $\mathbf{S_O}$ includes contributions from representation error, measurement error, and GEOS-Chem model error [*Heald et al.*, 2004]. Representation error is assumed to be negligible because SCIAMACHY X_{CH4} observations have horizontal footprints (30 km × 60–120 km) comparable to the size of GEOS-Chem grid cells. We use reported IMAP v5.5 values for the measurement error (standard deviation 30.2 ppb or 1.7%) since these are consistent with our INTEX-A validation (section 2). GEOS-Chem comparison to HIPPO vertical profiles across the Pacific indicates a model error standard deviation of 16 ppb for methane column mixing ratios, and we assume that this holds for North America too. All errors are assumed to be Gaussian and are added in quadrature to calculate the observational error for each observation. We do not include error correlation between observations since the overall observational error variance is dominated by the measurement error for which no correlation is found in the validation presented above.

The iterative optimization is implemented as follows. First, we perform five adjoint iterations to reduce $J(\mathbf{x_B})$. We then use the updated values of $\mathbf{x_B}$ to calculate $J(\mathbf{x_E})$ and perform five iterations to reduce $J(\mathbf{x_E})$. We use the updated values of $\mathbf{x_E}$ to recalculate $J(\mathbf{x_B})$ and repeat. When the reduction of the cost function at each iteration becomes small (0.5% of the cumulative cost function reduction up to that point), after 40 iterations, we hold $\mathbf{x_B}$ constant, and iteratively solve $\nabla_{x_E} J(\mathrm{x_E}) = 0$. Optimization of $\mathbf{x_B}$ corrects background methane for the inversion and is of peripheral interest here. We focus our discussion on the optimization of $\mathbf{x_E}$.

3.3. Clustering

Figure 5 shows the results from the inversion described above as optimized correction factors to the a priori methane emissions at $1/2° \times 2/3°$ horizontal resolution. Correction factors are weak, less than 30% for 93% of

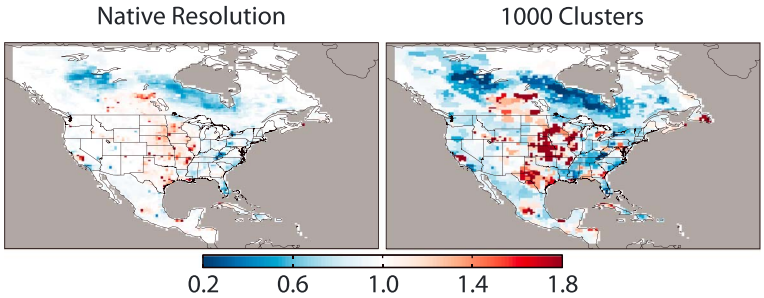

Figure 5. Emission scale factors relative to the a priori (Figure 3, top left) from inversions optimizing emissions (left) for the 1/2° × 2/3° native resolution of GEOS-Chem and (right) for 1000 clustered regions. Gray areas (ocean/ice) are not included in the state vector for the inversion.

grid cells. This is because the observations have insufficient information to constrain emissions at that resolution. As the discretization of emissions becomes finer, the observations become less sensitive to emissions from each grid cell. The inversion therefore has less ability to pull emissions in each grid cell away from their a priori value, and the optimal solution will be more tightly constrained by the a priori. This can be seen quantitatively from the minimization of (6):

$$\nabla_{\mathbf{x_E}} J(\mathbf{x_E}) = 2(\nabla_{\mathbf{x_E}} \mathbf{F})^{\mathbf{T}} \mathbf{S_O}^{-1}(F(\mathbf{x_E}) - \mathbf{y}) + 2\mathbf{S_{E,A}}^{-1}(\mathbf{x_E} - \mathbf{x_{E,A}}) = 0 \quad (8)$$

where $\nabla_{\mathbf{x_E}} \mathbf{F}$ is the Jacobian matrix of the forward model. As the dimension of $\mathbf{x_E}$ increases, the Jacobian matrix values become smaller, and thus, the individual terms of $(\nabla_{\mathbf{x_E}} \mathbf{F})^T \mathbf{S_O}^{-1}(F(\mathbf{x_E}) - \mathbf{y})$ decrease in magnitude as $(\nabla_{\mathbf{x_E}} \mathbf{F})^{\mathbf{T}}$ distributes $\mathbf{S_O}^{-1}(F(\mathbf{x_E}) - \mathbf{y})$ over a larger number of state vector elements. By contrast, the magnitude of individual terms of $\mathbf{S_{E,A}}^{-1}(\mathbf{x_E} - \mathbf{x_{E,A}})$ does not change. Thus, the a priori increases in importance relative to the observations.

The problem could be mitigated by accurately specifying error correlations in the a priori or by imposing them in the solution, as is done in geostatistical inversions [*Michalak et al*., 2004], but there is little confidence to be had in the specification of error correlations for methane sources. It could also be avoided altogether by optimizing grid cell fluxes rather than scaling factors (equation (7)) in the inversion, but this would require specification of absolute rather than relative errors for each grid cell.

We opted therefore to reduce the dimension of our emission state vector by clustering of grid cells, taking advantage of the results from the native resolution inversion (Figure 5) to group together neighboring grid cells with similar emission scale factors and thus minimize the aggregation error associated with clustering. We tried successively smaller numbers of clusters and repeated the inversion in the same manner described above for the native resolution inversion, seeking to find the best number of clusters for the inversion as measured by the fit to observations. As we initially decrease the number of clusters starting from the native resolution, we can expect an improved fit of the inversion results to the observations for the reasons discussed above. However, as the spatial resolution of the state vector becomes too coarse (too few clusters), the fit to observations degrades because of aggregation error.

We use a hierarchical clustering algorithm [*Johnson*, 1967] as a data-driven aggregation technique to optimally define clusters from the native resolution emissions grid. The algorithm initially assigns each 1/2° × 2/3° grid cell to its own region, calculates the "distance" to all other regions, and joins the two most similar. Distance is calculated as follows. We define the location for a region *I* by the vector $\mathbf{v_I} = (\mathbf{p}, 0.05\, s)^T$ where $\mathbf{p}$ is the location of the region centroid on a sphere and *s* is the mean value of the optimized scale factor from the native resolution inversion presented in Figure 5. All variables are normalized to unit variance and zero mean. The factor 0.05 was selected to adjust the weight of scale factors relative to geographic distance. The distance between two regions *I* and *m* is calculated as the norm $||\mathbf{v_I} - \mathbf{v_m}||$. The process of joining the two most similar regions proceeds iteratively, reducing the number of regions by one during each step. The algorithm can be stopped at any stage so that any number of clusters can be constructed.

Figure 6 (black) shows the contribution of the observation term, $(F(\hat{\mathbf{x}}_\mathbf{E}) - \mathbf{y})^T \mathbf{S_O}^{-1}(F(\hat{\mathbf{x}}_\mathbf{E}) - \mathbf{y})$, to the optimized cost function for inversions performed using different numbers of clustered regions. Here $\hat{\mathbf{x}}_\mathbf{E}$ is the

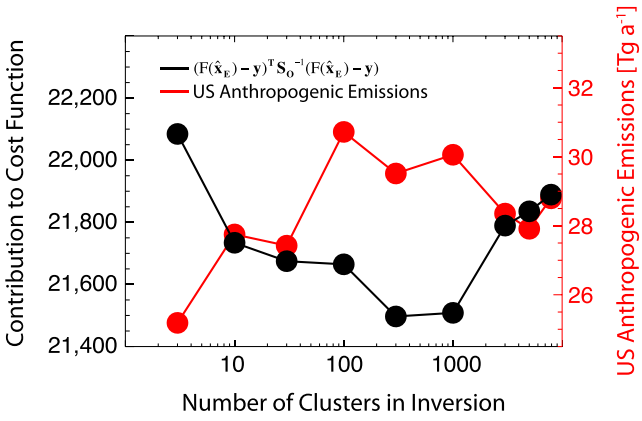

Figure 6. Sensitivity of inversion results to the resolution with which North American methane emissions are optimized from the SCIAMACHY data for 1 July to 14 August 2004. Resolution is expressed as the number of spatial clusters used in the inversion. The maximum of 7906 clusters represents the native 1/2° × 2/3° grid of GEOS-Chem. Optimal aggregation of grid cells based on proximity and emission correction tendencies yields successively smaller numbers of clusters. Black points show the observation term of the cost function $(F(\hat{\mathbf{x}}_\mathbf{E}) - \mathbf{y})^T \mathbf{S_O}^{-1} (F(\hat{\mathbf{x}}_\mathbf{E}) - \mathbf{y})$ describing the ability of the cost function to fit the SCIAMACHY observations. Red points show optimized U.S. anthropogenic emissions for each inversion.

optimal estimate from the inversion. We do not include the a priori term since it depends on the number of clusters used. The best results are achieved for 300–1000 clusters. As the number of clusters decreases from 7906 (native resolution) to 1000, the observations become more sensitive to elements in the state vector, producing a better model fit. As the number of clusters decreases below 300, aggregation error degrades the model fit. The range in the cost function for the different inversions is relatively small because the measurement error dominates for any individual data point. We use the inversion with 1000 clusters as our best estimate in terms of optimization and spatial detail. Figure 7 shows the 1000 clusters used in this analysis.

Figure 5 (right) shows the correction factors to the a priori methane emissions from the 1000 cluster inversion. Patterns are similar to the native resolution inversion (Figure 5, left), but correction factors are much larger, reflecting the stronger influence from the observations. Total U.S. anthropogenic emissions are only weakly sensitive to the number of clusters used. The variability of inversion results using different numbers of clusters will be used in section 4 to derive uncertainty estimates for our optimal emissions.

3.4. Evaluation With SCIAMACHY and INTEX-A Data

Figure 8 shows optimized emissions, calculated as the product of optimized correction factors and prior emissions in each grid cell. We checked for improvement of the model fit to the SCIAMACHY data by comparing GEOS-Chem simulations with optimized versus a priori emissions and boundary conditions. For this we calculated the GEOS-Chem-SCIAMACHY root-mean-square difference (RMSD) and correlation coefficient (R) for the ensemble of 1/2° × 2/3° grid cells with SCIAMACHY data, averaged over the 1 July to 14 August 2004 period and weighted by the number of SCIAMACHY observations in each grid cell. We find that the inversion reduces the model-observation RMSD from 11.6 to 9.7 ppb, while R increases from 0.65 to 0.76. This demonstrates improvement, limited by the random noise in the individual SCIAMACHY measurements.

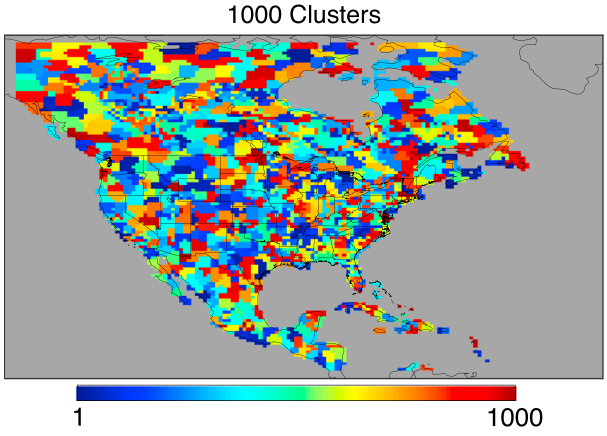

Figure 7. Map of 1000 clusters used in the optimal inversion. Colors represent a unique index number for each cluster.

We further used the boundary layer observations from INTEX-A (Figure 1) to provide verification of the inversion results. The model-observation RMSD for individual observations decreases from 33.5 to 28.5 ppb, while R increases slightly from 0.73 to 0.74. Here the improvement appears to be limited by small-scale model and representation error for individual observations. Averaging of the data allows us to reduce that error and is a more useful comparison. Figure 9 shows boundary layer (>850 hPa) GEOS-Chem-INTEX-A differences averaged on an 8° × 10° horizontal grid and for the INTEX-A period. The resulting model-observation RMSD

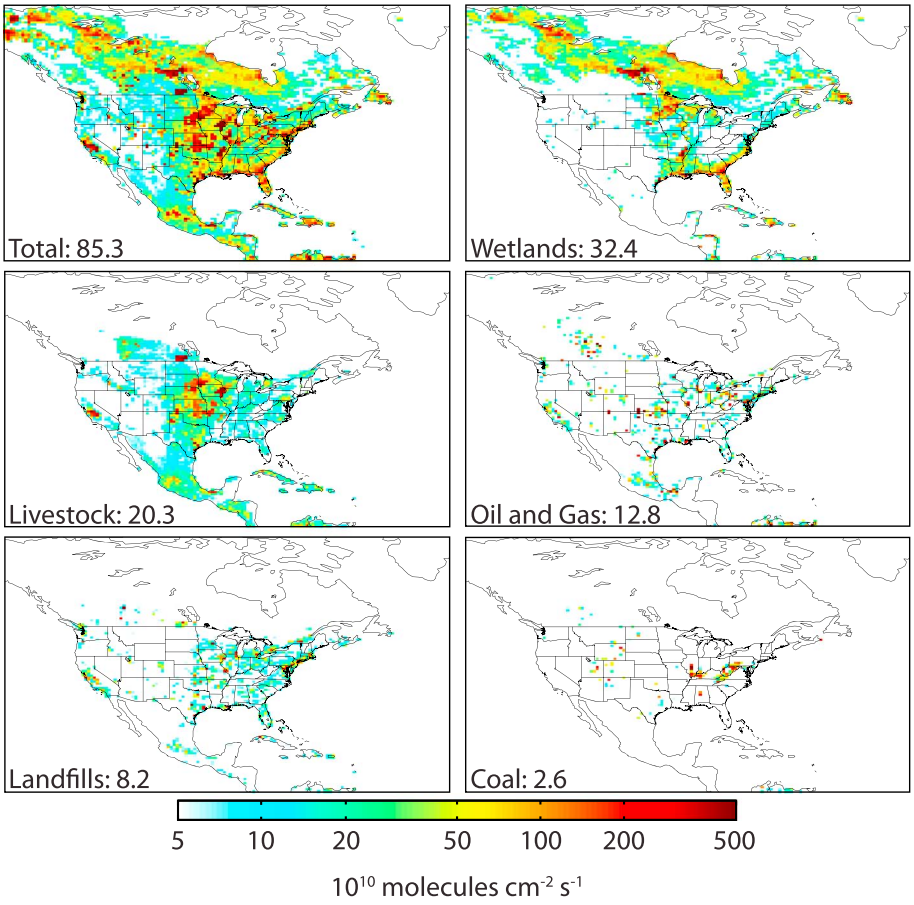

Figure 8. Optimized North American methane emissions from the 1000 cluster inversion: (top left) total emissions and contributions from the major source types. The annual emission rate for 2004 (Tg a^{-1}) is shown inset for the North America domain as encompassed by the figure.

weighted by the number of INTEX-A observations in each 8° × 10° grid cell decreases 23.2 to 12.3 ppb when using optimized instead of a priori emissions. The correlation coefficient *R* increases from 0.69 to 0.88.

We performed sensitivity inversions to investigate the effects of a priori constraints on emissions and model bias. A native resolution inversion without a priori constraints on emissions shows similar signs and patterns of emission corrections to the inversion with a priori constraints, but the magnitudes of corrections are larger. Evaluation using INTEX-A data averaged into 8° × 10° regions as above does not show as good a fit to observations, with an RMSD of 14.5 ppb and *R* of 0.77. This indicates that the a priori inventory contributes useful information. A sensitivity inversion including a uniform positive bias correction of 15 ppb in GEOS-Chem on the basis of INTEX-A free tropospheric data shows negligible effect on the correction factors to emissions because most of the bias is absorbed by correction to the boundary conditions.

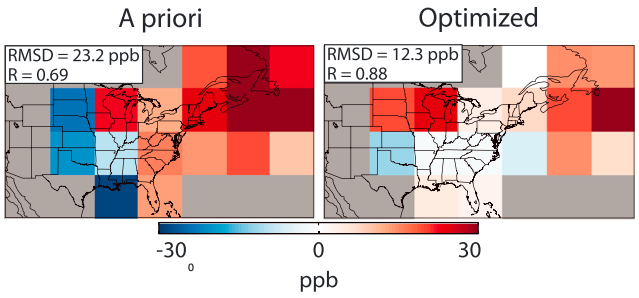

Figure 9. Evaluation of the SCIAMACHY inversion of methane emissions using INTEX-A aircraft data. The panels show the mean differences between GEOS-Chem and INTEX-A observations below 850 hPa and for 8° × 10° grid squares in the simulation (left) with a priori emissions and (right) with optimized emissions from the 1000 cluster inversion. A priori and optimized emission maps are shown in Figures 2 and 7. The model-observation root-mean-square difference and weighted correlation coefficient (*R*) are inset.

4. Optimized Methane Emissions

The optimized correction factors in Figure 5 show patterns of increases and decreases relative to the a priori emissions in Figure 3. There is a large decrease in emissions from natural wetlands in northern Ontario and western Canada, which drives a slight decrease in total North American emissions (85.3 Tg a^{-1} in the optimized emissions versus 90.2 Tg a^{-1} in EDGAR v4.2). There is also a broad decrease in the eastern U.S., particularly in Appalachia, suggesting an overestimate of emissions from coal mining and waste management. Emissions in the central U.S. and the Canadian Great Plains increase, suggesting an underestimate of emissions from livestock and also possibly from natural gas and oil extraction. Mexico City emissions increase. We elaborate below on source attribution for the U.S.

Table 1 shows U.S. anthropogenic emission estimates from EPA, EDGAR v4.2, and this work. The adjoint inversion method does not return uncertainties on optimized values, but we can estimate uncertainties from the ensemble of inversions with different numbers of clusters in Figure 6 (excluding the too coarse three-cluster inversion). Our optimized U.S. anthropogenic emissions are thus 30.1 ± 1.3 Tg a^{-1}, where the best estimate is from the 1000 cluster inversion. This is 17% higher than EDGAR v4.2 and 6% higher than *EPA* [2013] though within the stated EPA 95% confidence interval of 14%.

Our inversion optimizes the geographic distribution of emissions without a priori information on source type. It is of interest to determine whether the corrections can be attributed to the particular source types of Figure 3. To do so, we multiply the optimized emission correction factors by the a priori source estimates for each source type and grid cell. This approach assumes that the EDGAR v4.2 partition of source categories in each grid cell is correct. It does not assume that a priori spatial distributions of sources are correct because spatial patterns will change after applying the correction factors. Figure 8 shows the resulting spatial distributions of the five largest source types. Results in Table 1 include uncertainties from the inversions with different numbers of clusters as described above. These uncertainties are likely underestimates because they do not include contributions from errors in the a priori location of sources, in the CTM transport, and in the SCIAMACHY measurements.

To calculate annual emissions from our inversion results, we must assume knowledge of the seasonality of emissions. Most anthropogenic emissions are assumed to be aseasonal, as in the EDGAR v4.2 inventory. The *EPA* [2013] inventory, however, estimates livestock emissions during the summer (June, July, and August) to be 16% higher than the annual average, driven by temperature sensitivity of emissions from manure management. We account for this seasonality when estimating annual emission rates in this study. Uncharacterized seasonality in sources may contribute to overall uncertainty in extrapolating our inversion results to annual emissions, particularly from potential seasonality in distribution of natural gas and maintenance of natural gas systems.

Results in Table 1 show that our emission estimate for livestock is 40% higher than the EDGAR v4.2 and EPA inventories and represents the largest U.S. source. Our oil and gas source is intermediate between EDGAR v4.2 and EPA and is smaller than the livestock source. Other sources are also smaller, across all inventories. Our landfills source is consistent with both EDGAR v4.2 and EPA, while our coal mining source is smaller than EDGAR v4.2 but consistent with EPA.

The patterns of correction factors from our inversion in Figure 5 reveal structure that cannot be simply explained by the EDGAR v4.2 source types. A multiple-linear regression of absolute corrections on the distributions of individual a priori source types yields an R^2 of only 0.21. For example, our correction factors in Figure 5 indicate a large EDGAR underestimate of livestock emissions in parts of Iowa and southern Minnesota, where hog manure is important, but a decrease in eastern North Carolina where hog manure is important too. This could reflect differences in manure management practices. Our inversion also calls for a large increase in emissions from the Permian Basin in western Texas, a major oil and gas production region, but the EDGAR v4.2 inventory is very low there. This suggests that oil and gas emissions in EDGAR are too heavily weighted by the distribution and end use sectors relative to the production sector.

5. Comparison to Previous Studies

A number of previous studies have used methane observations from surface sites and aircraft as top-down constraints on methane emissions in North America. We discuss here the consistency of our results with these studies.

There has been much interest in quantifying wetland emissions from the Hudson Bay Lowlands (HBL) of northern Ontario, the second largest area of boreal wetlands in the world after western Siberia. *Pickett-Heaps et al.* [2011] reviewed previous studies and estimated an HBL source of 2.3 Tg a^{-1} from aircraft and surface observations. *Miller et al.* [2014] estimated a source of 2.4 Tg a^{-1} using tall tower observations. We find here a consistent estimate of 2.1 Tg a^{-1}, accounting for the seasonality given by *Pickett-Heaps et al.* [2011] with strong peak in June–August. This estimate may not be directly applicable to other years as HBL emissions exhibit significant interannual variability [*Zhu et al.*, 2013].

The California Research at the Nexus of Air Quality and Climate Change (CalNex) aircraft campaign in May–June 2010 provided constraints on methane emissions from California through a series of boundary layer flights across the state. Inverse analyses of the CalNex data by Santoni et al. (submitted manuscript, 2014) and *Wecht et al.* [2014] indicate statewide emissions of 2.4–2.8 Tg a^{-1}, Los Angeles Basin emissions of 0.3–0.4 Tg a^{-1}, and a factor of 2–4 underestimate of livestock emissions in the EDGAR v4.2 inventory for the Central Valley. Our inversion of the SCIAMACHY data is closely consistent with these results indicating a statewide emission of 2.1 Tg a^{-1} in California, 0.2 Tg a^{-1} in the Los Angeles Basin, and a factor of 2.6 underestimate in livestock emissions relative to EDGAR v4.2. The livestock underestimate is larger for California than the national underestimate of 40% reported earlier and provides further evidence of spatial errors in emission factors in the EDGAR inventory.

Miller et al. [2013] estimated methane emissions across the U.S. using a network of surface and aircraft data from 2007 to 2008. Their optimal estimate for U.S. anthropogenic emissions is 44.5 Tg a^{-1}, much higher than our value of 30.1 Tg a^{-1}. They increase emissions in the central U.S. relative to EDGAR and decrease emissions in Appalachia, similar to the spatial patterns reported here. Their observations, however, are relatively sparse east of the Great Plains and may not adequately characterize the emissions reductions throughout Appalachia, and the Northeast that are required by SCIAMACHY and consistent with the INTEX-A data. *Miller et al.* [2013] estimate a factor of 2.3 increase relative to EDGAR v4.2 for the northern plains (Nebraska, Iowa, Wisconsin, Minnesota, and South Dakota), a region of high livestock density and few other sources of methane. This compares well to our factor of 2.2 increase for the region, again higher than the national average for livestock emissions.

Katzenstein et al. [2003] measured methane concentrations on a road survey across Texas, Oklahoma, and Kansas. Assuming a mean boundary layer height and a characteristic ventilation time for the region, they estimated a methane emission of 4–6 Tg a^{-1} for that tristate region. Assuming that these emissions are mainly from natural gas and oil, they concluded that EPA emission estimates are too low by a factor of 2.5. Our inversion indicates a methane emission of 8.7 Tg a^{-1} for the region, greater than the *Katzenstein et al.* [2003] estimate and possibly reflecting their oversimplified ventilation model. In any case, their assumption that oil and gas dominate sources in the region may not be valid as the EDGAR v4.2 inventory for the region (4.0 Tg a^{-1}) assigns 52% of methane emissions to livestock and only 29% to oil and gas. *Miller et al.* [2013] estimate emissions of 10.8 Tg a^{-1} for the region and attribute the underestimate to both livestock and oil and gas. Most of the emission correction for the region in our inversion is from livestock.

Xiao et al. [2008] derived a U.S. ethane emission of 2.4 Tg a^{-1} from analysis of INTEX-A observations and combined this with independent estimates of ethane-to-methane emission ratios to deduce a U.S. fossil fuel methane emission of 16 Tg a^{-1}. This would include contributions from natural gas, oil, and coal mining. By comparison, EPA and EDGAR v4.2 estimate U.S. fossil fuel emissions of 11.7 and 10.1 Tg a^{-1}, respectively. Using the source-type distributions provided by EDGAR, we calculate a fossil fuel methane source of 9.6 Tg a^{-1}. Ethane-to-methane emission ratios are highly variable and may be a major uncertainty in the *Xiao et al.* [2008] estimate.

Kort et al. [2008] used Lagrangian modeling of observations from the North American CO_2 Budget and Regional Airborne Study (COBRA-NA) aircraft campaign across North America in 2003 to evaluate U.S. and Canada emissions. From a single-linear regression of modeled versus observed methane they estimated U.S. anthropogenic emissions of 41 ± 6 Tg a^{-1}, larger than our best estimate of 30.1 Tg a^{-1}. Their observations, however, are only sensitive to emissions from a relatively small fraction of the U.S. and Canada.

Pétron et al. [2012] and *Karion et al.* [2013] used in situ observations to estimate methane leak rates of 4% and 6–12% of total natural gas production Weld County, CO, and Uintah County, UT, respectively. In contrast, the EDGAR v4.2 and EPA inventories assume a national average leak rate of 1.0–1.4%. Our inversion of the

SCIAMACHY data does not indicate EDGAR v4.2 natural gas emissions to be underestimated from these two counties, but this would not account for post-2004 growth. Methane leakage rates can vary considerably by basin [*U.S. Government Accountability Office*, 2010].

Brandt et al. [2014] indicate that methane emissions from natural gas systems across the U.S. appear larger than estimated by the EPA. However, most studies cited by *Brandt et al.* [2014] focused on natural gas producing regions. While we do not find a significant continental-scale underestimate of emissions from natural gas, the spatial pattern of our correction factors suggests that a priori emissions are too heavily weighted by distribution relative to production.

6. Conclusions

We used SCIAMACHY satellite observations in a high-resolution continental-scale inversion of methane emissions in North America driven by the GEOS-Chem chemical transport model (CTM) and its adjoint at 1/2° × 2/3° horizontal resolution. The inversion focused on summer 2004, when concurrent observations from the INTEX-A aircraft observations are available to both validate the SCIAMACHY data and evaluate the inversion. The high density of observations available from SCIAMACHY enables finer spatial detail in constraining methane emissions on the continental scale than had been achievable by surface and aircraft observations.

Removal of measurement bias is essential for a successful inversion. Our validation of the SCIAMACHY observations (IMAP v5.5) with INTEX-A vertical profiles identified a systematic bias correlated with water vapor, consistent with previous studies. We found that we could successfully correct for this bias, and the residual error is consistent with the theoretical error estimate from the IMAP v5.5 retrieval.

Continental-scale inversion for methane required accurate specification of the boundary conditions in GEOS-Chem. This was accomplished by optimizing both the North American emissions and the boundary conditions as part of the inversion. We found that the information content from the SCIAMACHY data was insufficient to constrain emissions at the native 1/2° × 2/3° horizontal resolution of GEOS-Chem. We solved this problem by using a hierarchical clustering algorithm to identify 1000 geographical clusters for which the inversion provides optimal results. The optimized emissions obtained from the 1000 cluster inversion were evaluated by GEOS-Chem simulation of the INTEX-A aircraft data. This demonstrated a major improvement over the simulation driven by a priori emissions.

Our optimized methane emissions for the Canadian wetlands are lower than the a priori but consistent with recent studies. Our optimized methane emissions for the U.S. are lower than the EDGAR v4.2 inventory for the eastern U.S. but higher for the central U.S. Our best estimate of U.S. anthropogenic emissions is $30.1 \pm 1.3\ \mathrm{Tg\,a^{-1}}$, compared to $25.8\ \mathrm{Tg\,a^{-1}}$ and $28.3\ \mathrm{Tg\,a^{-1}}$ in the EDGAR v4.2 and EPA inventories, respectively. Source attribution of our optimized methane emissions on the basis of the EDGAR patterns suggests that the above inventories underestimate livestock emissions by 40%, with smaller discrepancies for other sources. We find that livestock emissions in the U.S. are 70% higher than oil and gas emissions, whereas the EPA inventory reports these two sources to be of comparable magnitude. However, we find in a regression analysis that the EDGAR patterns can account for only 21% of the variability in the source correction from the inversion. This implies large inventory errors in the geographic variability of emission factors (e.g., livestock management practices) and activity rates (e.g., oil and gas production in the west).

Our finding that U.S. livestock emissions are underestimated in current inventories is consistent with previous regional studies. The degree of underestimate seems highly variable for different parts of the country, suggesting large variability in emission factors. Our finding that oil and gas emissions are not underestimated in current inventories is at odds with previous studies and may partly reflect variability in leakage rates and assumptions made in all studies to perform source-type attribution.

Emissions of methane in North America may be rapidly changing in the future as a result of increasing oil and gas production, changes in recovery practices, evolving regulations, and climate change affecting wetlands. The GOSAT satellite observations (2009 to present) may be useful to track recent trends but are relatively sparse. The Tropospheric Monitoring Instrument (TROPOMI) instrument to be launched in 2015 will provide global daily coverage with $7 \times 7\ \mathrm{km^2}$ nadir spatial resolution and accuracy and precision of 2.0 and 0.6%, respectively [*Veefkind et al.*, 2012; *Butz et al.*, 2012]. This will provide a tremendous boost to monitoring methane emissions from space.

Acknowledgments
This work was supported by the NASA Carbon Monitoring System (CMS), the NASA Atmospheric Composition Modeling and Analysis Program (ACMAP), and by a NASA Earth System Science Fellowship to K.J.W. We thank Tom Wirth (U.S. EPA) for providing information on seasonality in the EPA emission inventory. The INTEX-A data are available through NASA's LaRC Airborne Science Data for Atmospheric Composition: ftp://ftp-air.larc.nasa.gov/pub/INTEXA/DC8_AIRCRAFT/. The SCIAMACHY data are available upon request through the SCIAMACHY website: http://www.sciamachy.org/products/index.php. Instructions for downloading and running the GEOS-Chem CTM are available at http://geos-chem.org/ and for the GEOS-Chem adjoint at http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_Adjoint.

References

Allen, D., K. Pickering, B. Duncan, and M. Damon (2010), Impact of lightning NO emissions on North American photochemistry as determined using the Global Modeling Initiative (GMI) model, *J. Geophys. Res.*, *115*, D22301, doi:10.1029/2010JD014062.

Bergamaschi, P., et al. (2007), Satellite chartography of atmospheric methane from SCIAMACHY on board ENVISAT: 2. Evaluation based on inverse model simulations, *J. Geophys. Res.*, *112*, D02304, doi:10.1029/2006JD007268.

Bergamaschi, P., et al. (2009), Inverse modeling of global and regional CH_4 emissions using SCIAMACHY satellite retrievals, *J. Geophys. Res.*, *114*, D22301, doi:10.1029/2009JD012287.

Bergamaschi, P., et al. (2013), Atmospheric CH_4 in the first decade of the 21st century: Inverse modeling analysis using SCIAMACHY satellite retrievals and NOAA surface measurements, *J. Geophys. Res. Atmos.*, *118*, 7350–7369, doi:10.1002/jgrd.50480.

Bocquet, M. (2005), Grid resolution dependence in the reconstruction of an atmospheric tracer source, *Nonlinear Processes Geophys.*, *12*, 219–234.

Bocquet, M. (2009), Toward optimal choices of control space representation for geophysical data assimilation, *Mon. Weather Rev.*, *137*, 2331–2348, doi:10.1175/2009MWR2789.1.

Bocquet, M., L. Wu, and F. Chevallier (2011), Bayesian design of control space for optimal assimilation of observations. Part I: Consistent multiscale formalism, *Q. J. R. Meteorol. Soc.*, *137*, 1340–1356, doi:10.1002/qj.837.

Brandt, A. R., et al. (2014), Methane leaks from North American natural gas systems, *Science*, *343*, 733–735, doi:10.1126/science.1247045.

Butz, A., A. Galli, O. Hasekamp, J. Landgraf, P. Tol, and I. Aben (2012), TROPOMI aboard Sentinel-5 Precursor: Prospective performance of CH_4 retrievals for aerosol and cirrus loaded atmospheres, *Remote Sens. Environ.*, *120*, 267–276, doi:10.1016/j.rse.2011.05.030.

Colman, J. J., A. L. Swanson, S. Meinardi, B. C. Sive, D. R. Blake, and F. S. Rowland (2001), Description of the analysis of a wide range of volatile organic compounds in whole air samples collected during PEM-Tropics A and B, *Anal. Chem.*, *73*, 3723–3731, doi:10.1021/ac010027g.

Considine, D. B., J. A. Logan, and M. A. Olsen (2008), Evaluation of near-tropopause ozone distributions in the Global Modeling Initiative combined stratosphere/troposphere model with ozonesonde data, *Atmos. Chem. Phys.*, *8*(9), 2365–2385, doi:10.5194/acp-8-2365-2008.

Cressot, C., et al. (2014), On the consistency between global and regional methane emissions inferred from SCIAMACHY, TANSO-FTS, IASI and surface measurements, *Atmos. Chem. Phys.*, *14*, 577–592, doi:10.5194/acp-14-577-2014.

Crevoisier, C., et al. (2013), The 2007–2011 evolution of tropical methane in the mid-troposphere as seen from space by MetOp-A/IASI, *Atmos. Chem. Phys.*, *13*, 4279–4289, doi:10.5194/acp-13-4279-2013.

Dils, B., et al. (2006), Comparisons between SCIAMACHY and ground-based FTIR data for total columns of CO, CH_4, CO_2 and N_2O, *Atmos. Chem. Phys.*, *6*, 1953–1976, doi:10.5194/acp-6-1953-2006.

European Commission, Joint Research Centre/Netherlands Environmental Assessment Agency (2009), Emission Database for Global Atmospheric Research (EDGAR), release version 4.0. [Available at http://edgar.jrc.ec.europa.eu.]

Frankenberg, C., J. F. Meirink, P. Bergamaschi, A. P. H. Goede, M. Heimann, S. Kröner, U. Platt, M. van Weele, and T. Wagner (2006), Satellite chartography of atmospheric methane from SCIAMACHY on board ENVISAT: Analysis of the years 2003 and 2004, *J. Geophys. Res.*, *111*, D07303, doi:10.1029/2005JD006235.

Frankenberg, C., P. Bergamaschi, A. Butz, S. Houweling, J. F. Meirink, J. Notholt, A. K. Petersen, H. Schrijver, T. Warneke, and I. Aben (2008), Tropical methane emissions: A revised view from SCIAMACHY onboard ENVISAT, *Geophys. Res. Lett.*, *35*, L15811, doi:10.1029/2008GL034300.

Frankenberg, C., I. Aben, P. Bergamaschi, E. J. Dlugokencky, R. van Hees, S. Houweling, P. van der Meer, R. Snel, and P. Tol (2011), Global column-averaged methane mixing ratios from 2003 to 2009 as derived from SCIAMACHY: Trends and variability, *J. Geophys. Res.*, *116*, D02304, doi:10.1029/2010JD014849.

Fung, I., J. John, J. Lerner, E. Matthews, M. Prather, L. P. Steele, and P. J. Fraser (1991), Three-dimensional model synthesis of the global methane cycle, *J. Geophys. Res.*, *96*, 13,033–13,065, doi:10.1029/91JD01247.

Hartmann, D. L., et al. (2013), Observations: Atmosphere and surface, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., pp. 159–254, Cambridge Univ. Press, Cambridge, U. K., and New York.

Heald, C., D. Jacob, D. Jones, P. Palmer, J. Logan, D. Streets, G. Sachse, J. Gille, R. Hoffman, and T. Nehrkorn (2004), Comparative inverse analysis of satellite (MOPITT) and aircraft (TRACE-P) observations to estimate Asian sources of carbon monoxide, *J. Geophys. Res.*, *109*, D23306, doi:10.1029/2004JD005185.

Henze, D. K., A. Hakami, and J. H. Seinfeld (2007), Development of the adjoint of GEOS-Chem, *Atmos. Chem. Phys.*, *7*, 2413–2433.

Houweling, S., et al. (2013), A multi-year methane inversion using SCIAMACHY, accounting for systematic errors using TCCON measurements, *Atmos. Chem. Phys. Discuss.*, *13*, 28,117–28,171, doi:10.5194/acpd-13-28117-2013.

International Energy Agency (2013), *World Energy Outlook Special Report: Redrawing the Energy-Climate Map*, International Energy Agency, Paris, France, 10 June. [Available at http://www.worldenergyoutlook.org/media/weowebsite/2013/energyclimatemap/RedrawingEnergyClimateMap.pdf.]

Johnson, S. (1967), Hierarchical clustering schemes, *Psychometrika*, *32*(3), 241–254, doi:10.1007/BF02289588.

Kaplan, J. O. (2002), Wetlands at the Last Glacial Maximum: Distribution and methane emissions, *Geophys. Res. Lett.*, *29*(6), 1079, doi:10.1029/2001GL013366.

Karion, A., et al. (2013), Methane emissions estimate from airborne measurements over a western United States natural gas field, *Geophys. Res. Lett.*, *40*, 4393–4397, doi:10.1002/grl.50811.

Katzenstein, A. S., L. A. Doezema, I. J. Simpson, D. R. Blake, and F. S. Rowland (2003), Extensive regional atmospheric hydrocarbon pollution in the southwestern United States, *Proc. Natl. Acad. Sci. U.S.A.*, *100*(21), 11,975–11,979, doi:10.1073/pnas.1635258100.

Kirschke, S., et al. (2013), Three decades of methane sources and sinks: Budgets and variations, *Nat. Geosci.*, *6*, 813–823, doi:10.1038/ngeo1955.

Kort, E. A., J. Eluszkiewicz, B. B. Stephens, J. B. Miller, C. Gerbig, T. Nehrkorn, B. C. Daube, J. O. Kaplan, S. Houweling, and S. C. Wofsy (2008), Emissions of CH_4 and N_2O over the United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations, *Geophys. Res. Lett.*, *35*, L18808, doi:10.1029/2008GL034031.

Meirink, J. F., et al. (2008), Four-dimensional variational data assimilation for inverse modeling of atmospheric methane emissions: Analysis of SCIAMACHY observations, *J. Geophys. Res.*, *113*, D17301, doi:10.1029/2007JD009740.

Michalak, A. M., L. Bruhwiler, and P. P. Tans (2004), A geostatistical approach to surface flux estimation of atmospheric trace gases, *J. Geophys. Res.*, *109*, D14109, doi:10.1029/2003JD004422.

Miller, S. M., et al. (2013), Anthropogenic emissions of methane in the United States, *Proc. Natl. Acad. Sci. U.S.A.*, doi:10.1073/pnas.1314392110.

Miller, S. M., et al. (2014), Observational constraints on the distribution, seasonality, and environmental predictors of North American boreal methane emissions, *Global Biogeochem. Cycles*, *28*, 146–160, doi:10.1002/2013GB004580.

Monteil, G., S. Houweling, A. Butz, S. Guerlet, D. Schepers, O. Hasekamp, C. Frankenberg, R. Scheepmaker, I. Aben, and T. Röckmann (2013), Comparison of CH_4 inversions based on 15 months of GOSAT and SCIAMACHY observations, *J. Geophys. Res. Atmos.*, *118*, 11,807–11,823, doi:10.1002/2013JD019760.

Mu, M., et al. (2010), Daily and 3-hourly variability in global fire emissions and consequences for atmospheric model predictions of carbon monoxide, *J. Geophys. Res.*, *116*, D24303, doi:10.1029/2011JD016245.

Murray, L. T., D. J. Jacob, J. A. Logan, R. C. Hudman, and W. J. Koshak (2012), Optimized regional and interannual variability of lightning in a global chemical transport model constrained by LIS/OTD satellite data, *J. Geophys. Res.*, *117*, D20307, doi:10.1029/2012JD017934.

Myhre, G., et al. (2013), Anthropogenic and natural radiative forcing, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., pp. 659–740, Cambridge Univ. Press, Cambridge, U. K., and New York.

Park, R. J., D. J. Jacob, B. D. Field, R. M. Yantosca, and M. Chin (2004), Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: Implications for policy, *J. Geophys. Res.*, *109*, D15204, doi:10.1029/2003JD004473.

Park, R. J., D. J. Jacob, N. Kumar, and R. M. Yantosca (2006), Regional visibility statistics in the United States: Natural and transboundary pollution influences, and implications for the Regional Haze Rule, *Atmos. Environ.*, *40*(28), 5405–5423, doi:10.1016/j.atmosenv.2006.04.059.

Parker, R., et al. (2011), Methane observations from the Greenhouse Gases Observing SATellite: Comparison to ground-based TCCON data and model calculations, *Geophys. Res. Lett.*, *38*, L15807, doi:10.1029/2011GL047871.

Pétron, G., et al. (2012), Hydrocarbon emissions characterization in the Colorado Front Range: A pilot study, *J. Geophys. Res.*, *117*, D04304, doi:10.1029/2011JD016360.

Pickett-Heaps, C. A., D. J. Jacob, K. J. Wecht, E. A. Kort, S. C. Wofsy, G. S. Diskin, D. E. J. Worthy, J. O. Kaplan, I. Bey, and J. Drevet (2011), Magnitude of seasonality of wetland methane emissions from the Hudson Bay Lowlands (Canada), *Atmos. Chem. Phys.*, *11*(8), 3773–3779, doi:10.5194/acp-11-3773-2011.

Prather, M. J., C. D. Holmes, and J. Hsu (2012), Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry, *Geophys. Res. Lett.*, *39*, L09803, doi:10.1029/2012GL051440.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding*, World Scientific Co. Pte. Ltd., Tokyo.

Schepers, D., et al. (2012), Methane retrievals from Greenhouse Gases Observing Satellite (GOSAT) shortwave infrared measurements: Performance comparison of proxy and physics retrieval algorithms, *J. Geophys. Res.*, *117*, D10307, doi:10.1029/2012JD017549.

Singh, H. B., W. H. Brune, J. H. Crawford, and D. J. Jacob (2006), Overview of the summer 2004 Intercontinental Chemical Transport Experiment-North America (INTEX-A), *J. Geophys. Res.*, *111*, D24S01, doi:10.1029/2006JD007905.

Turner, A. J., D. J. Jacob, K. J. Wecht, M. Sulprizio, V. Payne, G. Santoni, S. C. Wofsy, K. W. Bowman, R. Parker, and H. Boesch (2013), Optimal estimation of North American methane emissions using GOSAT data: A contribution to the NASA Carbon Monitoring System, paper presented at 2013 Fall Meeting, AGU, San Francisco, Calif.

United States Environmental Protection Agency (EPA) (2013), Inventory of U.S. greenhouse gas emissions and sinks: 1990–2011-Annexes. [Available at http://www.epa.gov/climatechange/Downloads/ghgemissions/US-GHG-Inventory-2013-Main-Text.pdf.]

United States Government Accountability Office (2010), Federal Oil and Gas Leases: Opportunities exist to capture vented and flared natural gas, which would increase royalty payments and reduce greenhoues gases, Report to Congressional Requesters, Washington, D. C. [Available at http://www.gao.gov/assets/320/311826.pdf.]

van der Werf, G. R., J. T. Randerson, L. Giglio, G. J. Collatz, M. Mu, P. S. Kasibhatla, D. C. Morton, R. S. DeFries, Y. Jin, and T. T. van Leeuwen (2010), Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), *Atmos. Chem. Phys.*, *10*, 11,707–11,735, doi:10.5194/acp-10-11707-2010.

Van Donkelaar, A., R. V. Martin, A. N. Pasch, J. J. Szykman, L. Zhang, Y. X. Wang, and D. Chen (2012), Improving the accuracy of daily satellite-derived ground-level fine aerosol concentration estimates for North America, *Environ. Sci. Technol.*, *46*(21), 11,971–11,978, doi:10.1021/es3025319.

van Vuuren, D. P., E. Stehfest, M. G. J. den Elzen, J. van Vliet, and M. Isaac (2010), Exploring IMAGE model scenarios that keep greenhouse gas radiative forcing below 3 W/m^2 in 2100, *Energy Econ.*, *32*, 1105–1120.

Veefkind, J. P., et al. (2012), TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, *Remote Sens. Environ.*, *120*, 70–83, doi:10.1016/j.rse.2011.09.027.

Voulgarakis, A., et al. (2013), Analysis of present day and future OH and methane lifetime in the ACCMIP simulations, *Atmos. Chem. Phys.*, *13*, 2563–2587, doi:10.5194/acp-13-2563-2013.

Wang, J. S., J. L. Logan, M. B. McElroy, B. N. Duncan, I. A. Megretskaia, and R. M. Yantosca (2004), 3-D model analysis of the slowdown and interannual variability in the methane growth rate from 1988 to 1997, *Global Biogeochem. Cycles*, *18*, GB3011, doi:10.1029/2003GB002180.

Wecht, K. J., D. J. Jacob, S. C. Wofsy, E. A. Kort, J. R. Worden, S. S. Kulawik, D. K. Henze, M. Kopacz, and V. H. Payne (2012), Validation of TES methane with HIPPO aircraft observations: Implications for inverse modeling of methane sources, *Atmos. Chem. Phys.*, *12*, 1823–1832, doi:10.5194/acp-12-1823-2012.

Wecht, K. J., D. J. Jacob, M. P. Sulprizio, G. W. Santoni, S. C. Wofsy, R. Parker, and J. R. Worden (2014), Spatially resolving methane emissions in California: Constraints from the CalNex aircraft campaign and from present (GOSAT, TES) and future (TROPOMI, GEO-CAPE) satellite observations, *Atmos. Chem. Phys. Discuss.*, *14*, 4119–4148, doi:10.5194/acpd-14-4119-2014.

Weyant, J. P., F. C. de la Chesnaye, and G. J. Blanford (2006), Overview of EMF-21: Multigas mitigation climate policy, *Energy J.*, Multi-Greenhouse Gas Mitigation and Climate Policy Special Issue, *27*(3), 1–32.

Wofsy, S. C., and the HIPPO Science Team and Cooperating Modelers and Satellite Teams (2011), HIAPER Pole-to-Pole Observations (HIPPO): Fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols, *Philos. Trans. R. Soc. A*, *369*(1943), 2073–2086, doi:10.1098/rsta.2010.0313.

Worden, J., S. Kulawik, C. Frankenberg, K. Bowman, V. Payne, K. Cady-Peirara, K. Wecht, J.-E. Lee, D. Noone, and C. Risi (2012), Profiles of CH_4, HDO, H_2O, and N_2O with improved lower tropospheric vertical resolution from Aura TES radiances, *Atmos. Meas. Tech.*, *5*, 397–411, doi:10.5194/amt-5-397-2012.

Wu, L., M. Bocquet, T. Lauvaux, F. Chevallier, P. Rayner, and K. Davis (2011), Optimal representation of source-sink fluxes for mesoscale carbon dioxide inversion with synthetic data, *J. Geophys. Res.*, *116*, D21304, doi:10.1029/2011JD016198.

Xiao, Y., J. A. Logan, D. J. Jacob, R. C. Hudman, R. Yantosca, and D. R. Blake (2008), Global budget of ethane and regional constraints on U.S. sources, *J. Geophys. Res.*, *113*, D21306, doi:10.1029/2007JD009415.

Xiong, X., C. Barnet, E. Maddy, C. Sweeney, X. Liu, L. Zhou, and M. Goldberg (2008), Characterization and validation of methane products from the Atmospheric Infrared Sounder (AIRS), *J. Geophys. Res.*, *113*, G00A01, doi:10.1029/2007JD000500.

Xiong, X., C. Barnet, E. S. Maddy, A. Gambacorta, T. S. King, and S. C. Wofsy (2013), Mid-upper tropospheric methane retrieval from IASI and its validation, *Atmos. Meas. Tech.*, *6*, 2255–2265, doi:10.5194/amt-6-2255-2013.

Zhang, L., D. J. Jacob, N. V. Smith-Downey, D. A. Wood, D. Blewitt, C. C. Carouge, A. van Donkelaar, D. B. A. Jones, L. T. Murray, and Y. Wang (2011), Improved estimate of the policy-relevant background ozone in the United States using the GEOS-Chem global model with 1/2° × 2/3° horizontal resolution over North America, *Atmos. Environ.*, *45*, 6769–6776, doi:10.1016/j.atmosenv.2011.07.054.

Zhang, L., D. J. Jacob, E. M. Knipping, N. Kumar, J. W. Munger, C. C. Carouge, A. van Donkelaar, Y. Wang, and C. Chen (2012), Nitrogen deposition to the United States: Distribution, sources, and processes, *Atmos. Chem. Phys.*, *12*, 4539–44,554, doi:10.5194/acp-12-4539-2012.

Zhang, Y., et al. (2012), Nested-grid simulation of mercury over North America, *Atmos. Chem. Phys.*, *12*, 6095–6111, doi:10.5194/acp-12-6095-2012.

Zhu, X., et al. (2013), Estimating wetland methane emissions from the northern high latitudes from 1990 to 2009 using artificial neural networks, *Global Biogeochem. Cycles*, *27*, 592–604, doi:10.1002/gbc.20052.